"TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data" 👩🏻‍💻

-**Abstract** :

Biology has entered the big-data age thanks to the rapid advancement of high-throughput sequencing techniques. Programming and command-line environments are needed for data analysis using various bioinformatics tools, which are difficult and time-consuming for most wet-lab biologists. We present TBtools (a Toolkit for Biologists that integrates various biological data-handling tools) in this paper: TBtools can be used to create a standalone software with a user-friendly interface and a wide range of graphs using a new plotting engine ("JIGplot") built to optimize their interactive ability; this engine allows fast point-and-click adjustment of almost any graphic function. The toolkit includes over 130 functions ranging from bulk sequence processing to interactive data visualization. TBtools is platform-agnostic software that runs on any operating system that supports Java Runtime Environment 1.6 or higher. It is open to non-commercial users for free at **@https://github.com/CJ-Chen/TBtools/releases** .

-Introduction:

Wet-lab biologists face a huge challenge in effectively managing big data and interpreting the underlying bioinformation. To meet this challenge, a variety of bioinformatics software programmes , pipelines, and packages have been developed; however, the majority of these tools are packaged as scripts written in a variety of programming languages and require some command-line experience. For wet-lab biologists who want to process their own data but lack proficient analytical skills, the lack of easily accessible resources remains a major barrier. On a genomic scale, HTS technologies are commonly used to examine biological phenomena. Unfortunately, when experimental biologists run into programming roadblocks, the big data produced is often underutilised .We frequently run into problems when dealing with large data files created by HTS. For example, extracting specific sequence records or regional sequences from large Fasta files is a common task that is difficult or impossible to accomplish with traditional text editors, particularly in batch mode. While some scripts or applications, such as seqkit and FasPaser2, are available, they either require users to have a working knowledge of the command-line environment or are limited to Windows operating systems. A heatmap is a popular visual representation of large-scale gene expression data. Various web-based tools, such as Heatmapper , enable users to create heatmaps outside of the command line environment; however, these programmes require a stable internet connection, which prevents offline data analysis. Gene families are collections of genes with identical sequences and functions. Not only sequence features, but also gene structure, must be identified for comparative analyses ,conserved motifs, and potential functional domains, are essential in revealing the conservation and differences among members of a gene family. While tools like GSDS2 and FancyGenes have been developed to display gene structures, there is no single tool that can show all of these features at the same time. Furthermore, several genius methods, such as Upsetplots, Circos plots , and eFP browsers, have been created in response to the growing demand for viewing broad datasets . Each of these visualisation tools is bundled in a variety of programming environments, making them difficult to navigate for the typical wet-lab biologist. To make it even more difficult, many of these applications have a long list of dependencies, making software installation difficult. Circos, for example, comes with a long list of dependencies on other Perl modules, rendering the programme installation daunting for the average wetlab biologist. The majority of bioinformatics tools required for analysing and visualising large biological datasets lack user-friendly interfaces that allow for easy point-and-click interactions as well as interactive modification options. We built and present here TBtools to overcome this deficiency and bridge the gap between bioinformatics analyses and wet-lab data producers. TBtools is a cross-platform compatible integrative and collaborative toolkit. Our ultimate goal is to speed up discoveries by offering an out-of-the-box solution to biologists' data-handling conundrum.